

Protein network analysis

 Miguel A. Gonzalez-Lozano

Updated date: May 25, 2020



An abbreviated version of this protocol was published in Science Advances in Feb 2020

Stitching the synapse: Cross-linking mass spectrometry into resolving synaptic protein interactions

DOI: 10.1126/sciadv.aax5783

Detailed protocol

Cross-linking protein-protein interaction (PPI) networks can be generated based on the interprotein cross-links identified. Since multiple tools were used for the protein network analysis, in this protocol we indicate the tools that were utilized in each step and the references for further explanations.

1. Network creation:
 - Import the interprotein cross-linking data into the R language for statistical computation (1).
 - Generate a protein network using the *igraph* R package (2).
 - Cluster the proteins from the main component of the network by applying unsupervised edge-betweenness clustering (3) with the same package.
2. Functional annotation:
 - Annotate each cluster by performing Gene Ontology enrichment analysis using PANTHER classification system (4), with all the proteins identified as background.
 - Disconnected modules can be annotated individually or in groups defined by using DAVID Gene Functional Classification (5).
3. Additional protein and PPI data: further information from publicly available databases can be included for subsequent analysis and/or visualization. E.g.:
 - Protein domain information for the cross-linked sites can be retrieved from UniProt.
 - Protein interactions from literature can be obtained from STRINGdb (6), BioGRID (7) and InWEB (8).
4. Network analysis: add the functional annotations to the network and analyze using *igraph* R package.
 - Calculate the modularity score and degree distribution of the network.
 - Measure the network path distances and the number of protein pairs directly connected with the same GO annotation.
 - As control for the different measurements, the same package allows the random rewiring of the network while preserving the degree distribution using.
 - The results can be exported in multiple formats, such as table format and graphML.
5. Network visualization: the protein network can be visualized by using open-source software, such as Cytoscape (9) (<https://cytoscape.org/>) or Gephi (<https://gephi.org/>). While *igraph* package includes the visualization of networks, these software offers a user-friendly and quick manner to adjust multiple options of the network display. E.g.:
 - Adjust the layout of the network (e.g. organic, circular) and node position to avoid protein label overlap.
 - Display protein features by using node properties, such as nodes color based on the functional clusters and nodes size proportional to the protein abundance.
 - Display PPI features by using edge properties, such as edges color to indicate PPIs reported in the literature or sample of identification, and edges width proportional to the number of cross-links identified.

Cytoscape also allows the generation of the PPI network and part of the analyses indicated previously, representing a user-friendly alternative for users not familiar with R programming language.

References:

1. R. C. Team, R: A language and environment for statistical computing. *R Foundation for Statistical Computing*.
2. G. Csardi, T. Nepusz, The igraph software package for complex network research. *InterJournal Complex Syst.* **1695** (2006).
3. M. E. J. Newman, M. Girvan, Finding and evaluating community structure in networks. *Phys. Rev. E* **69**, 026113 (2004).
4. H. Mi *et al.*, Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). *Nat. Protoc.* **14**, 703–721 (2019).
5. D. W. Huang, B. T. Sherman, R. A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
6. D. Szklarczyk *et al.*, STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43**, D447–D452 (2015).
7. C. Stark *et al.*, BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* **34**, D535–9 (2006).
8. T. Li *et al.*, A scored human protein-protein interaction network to catalyze genomic interpretation. *Nat. Methods.* **14**, 61–64 (2017).
9. P. Shannon, Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **13**, 2498–2504 (2003).

How to cite: (Readers should cite both the Bio-protocol preprint and the original research article where this protocol was used)

1. Gonzalez-Lozano, M. A. (2020). Protein network analysis. Bio-protocol Preprint. bio-protocol.org/prep327.
2. Gonzalez-Lozano, M. A., Koopmans, F., Sullivan, P. F., Protze, J., Krause, G., Verhage, M., Li, K. W., Liu, F. and Smit, A. B. (2020). Stitching the synapse: Cross-linking mass spectrometry into resolving synaptic protein interactions. *Science Advances* 6(8). DOI: [10.1126/sciadv.aax5783](https://doi.org/10.1126/sciadv.aax5783)

Copyright: Content may be subjected to copyright.